text-torch

October 15, 2023

0.1 Text Torch - Information Retrieval from Documents

0.2 Objective

With the rise of Generative AI models these days, there has been an increase in the demand for document retrieval systems.

The objective of our system is to design and implement a supervised learning document retrieval system that efficiently finds the relevant information from the documents in response to user queries while ensuring the retrieval system provides results in a timely manner.

The system aims to capture the semantic context of the user's input and retrieve documents that are not only keyword-matching but also contextually relevant.

0.3 Use Case

We will be designing a General Purpose Document Retrieval System which can benefit a wide range of users.

- Researchers and Academicians (research papers, scholarly documents)
- Professionals (Be it legal, medical, finance or corporate sectors)
- Students and Educators

Moreover, this also aims for the individuals seeking information on various topics like healthcare, technology, cooking, or any domain can retrieve relevant documents, guides, and articles improving their knowledge and skills.

0.4 Data

As a baseline, we want our dataset to be generic and contains a vast spectrum of knowledge so that the system works well with wide range of users.

- We will be using Simple English Wikipedia Dataset from Published by SBert - https://www.sbert.net/, https://sbert.net/datasets
- Compared to the full English wikipedia, it has smaller number of articles and the amount of texts.

[ ]:
```
!pip install sentence_transformers
```

0.5 Preparing the Dataset

We will load json line file and convert it into a data frame.

[3]:
```
import os
import json
import pandas as pd
import numpy as np

import matplotlib.pyplot as plt
import seaborn as sns

from nltk.tokenize import word_tokenize
from tqdm import tqdm
from sklearn.metrics import jaccard_score
```

```
[nltk_data] Downloading package punkt to /root/nltk_data…
[nltk_data]   Unzipping tokenizers/punkt.zip.
```

[4]:
```
file_path = '/content/drive/MyDrive/datasets/simplewiki-2020-11-01.jsonl'

def parse_json(s):
  try:
      return json.loads(s)
  except:
      return None

datapoints = []
with open(file_path) as f:
    lines = f.read().splitlines()

    for line in lines:
      data = parse_json(line.strip())

      # if the JSON entry is corrupted, we will drop it
      if data == None:
        continue

      for paragraph in data['paragraphs']:
          # we encode the passages as {title, text}
          datapoints.append({
              'title': data['title'],
              'paragraph':  paragraph
          })

df = pd.DataFrame(datapoints)
```

0.6 Data Cleaning

[5]:
```
df.head(10)
```

[5]:

	title	paragraph
0	Ted Cassidy	Ted Cassidy (July 31, 1932 - January 16, 1979)…
1	Aileen Wuornos	Aileen Carol Wuornos Pralle (born Aileen Carol…
2	Aileen Wuornos	Wuornos was diagnosed with antisocial personal…
3	Aileen Wuornos	The movie, "Monster" is about her life. Two do…
4	Aileen Wuornos	Wuornos was born Aileen Carol Pittman in Roche…
5	Crater	A crater is a round dent on a planet. They are…
6	Store	Store has several meanings:
7	Chinese New Year	Chinese New Year, known in China as the Spring…
8	Chinese New Year	The Chinese New Year is of the most important …
9	Chinese New Year	Chinese New Year used to last 15 days until th…

[6]:
```
print(f'total number of estimated samples {df.shape[0]}\ntotal number of␣
 ↪features {df.shape[1]}')
```

```
total number of estimated samples 509663
total number of features 2
```

The dataset contains two features that we built up earlier and 500K samples which is a pretty good amount of data to demonstrate the document retrieval system.

[7]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 509663 entries, 0 to 509662
Data columns (total 2 columns):
 #   Column     Non-Null Count   Dtype
---  ------     --------------   -----
 0   title      509663 non-null  object
 1   paragraph  509663 non-null  object
dtypes: object(2)
memory usage: 7.8+ MB
```

0.6.1 Handling missing values

Drop NA

[8]:
```
total_values = df.size
missing_values = df.isna().sum()

missing_percentage = (missing_values / total_values) * 100

print("Missing values (na) count:")
print(missing_values)
```

```
print("\nPercentage of missing values (na):")
print(missing_percentage)
```

```
Missing values (na) count:
title        0
paragraph    0
dtype: int64

Percentage of missing values (na):
title        0.0
paragraph    0.0
dtype: float64
```

[9]:
```
df = df.dropna()
```

Drop value with empty strings We don't have much features. But it is important that our title and paragraph are not empty.

[10]:
```
# create a boolean mask for non-empty strings
non_empty_mask = (df['title'] != '') & (df['paragraph'] != '')

# filter the DataFrame using the mask
df = df[non_empty_mask]
```

Building a new Feature We will build a new feature to use across the EDA process.

[11]:
```
df['full_text'] = df['title'] + '\n' + df['paragraph']
```

0.7 Exploratory Data Analysis (EDA)

0.7.1 Document Length Analysis

We will first analyze the distribution of the token count of each passage title + paragraph.

[12]:
```
def tokenize_text(text):
    """
    Tokenizes input text using NLTK's word_tokenize function.

    Parameters:
    - text (str): The text to tokenize.

    Returns:
    - list of str: The tokenized text.
    """

    # tokenize text
    tokens = word_tokenize(text)
```

```
    return tokens
```

[13]:
```
df['token_count'] = df['full_text'].apply(lambda x: len(x.split()))
```

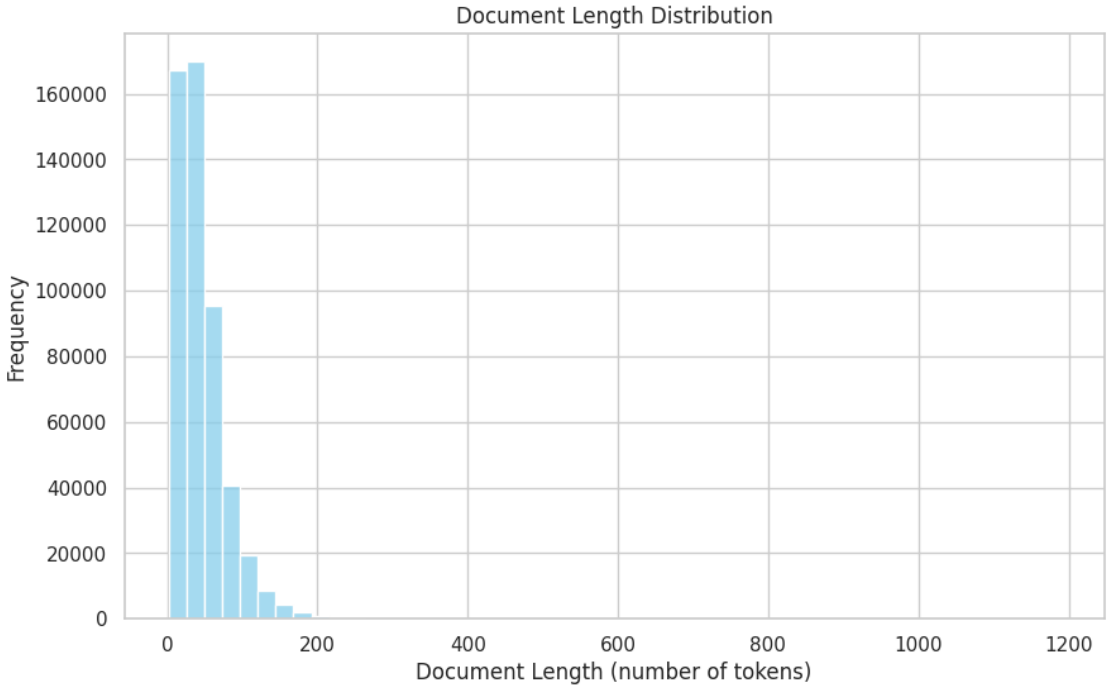
[14]:
```
mean_length = df['token_count'].mean()
median_length = df['token_count'].median()
mode_length = df['token_count'].mode()[0]  # mode() returns a Series, get the␣
↪first element
min_length = df['token_count'].min()
max_length = df['token_count'].max()

print(f"Mean length: {mean_length:.2f}")
print(f"Median length: {median_length}")
print(f"Mode length: {mode_length}")
print(f"Min length: {min_length}")
print(f"Max length: {max_length}")
```

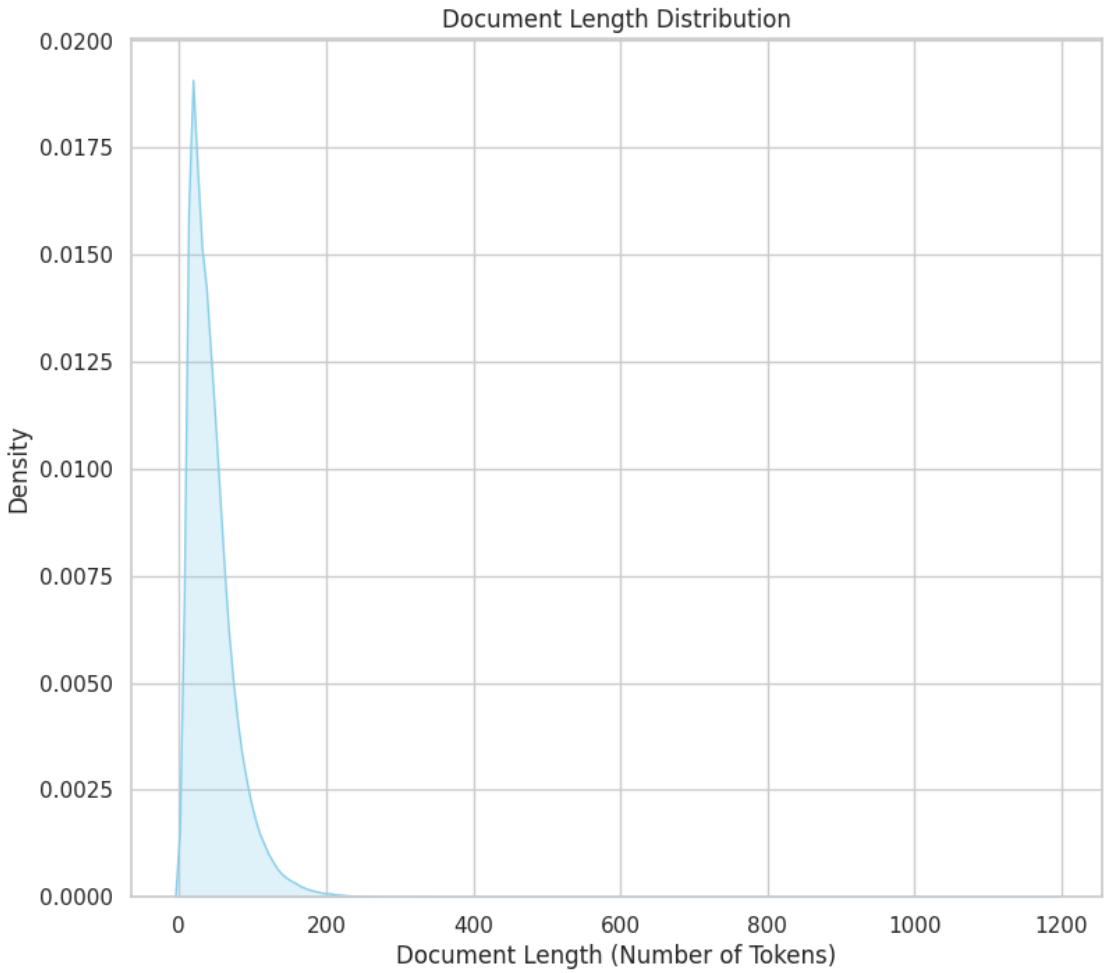
```
Mean length: 44.79
Median length: 37.0
Mode length: 21
Min length: 2
Max length: 1188
```

[15]:
```
sns.set(style="whitegrid")

# Histogram
plt.figure(figsize=(10, 6))
sns.histplot(df['token_count'], bins=50, kde=False, color='skyblue')
plt.title('Document Length Distribution')
plt.xlabel('Document Length (number of tokens)')
plt.ylabel('Frequency')
plt.show()
```



[16]:
```
# KDE plot
plt.figure(figsize=(9, 8))
sns.kdeplot(df['token_count'], color='skyblue', fill=True)
plt.title('Document Length Distribution')
plt.xlabel('Document Length (Number of Tokens)')
plt.ylabel('Density')
plt.show()
```



From our Histogram and KDE diagrams, we find out that

- Histogram: Shows a right-skewed distribution with most Wikipedia documents having a low token count.
- KDE Peak: Indicates the most common document length, reaffirming that most documents are short.
- Right Skew: Both plots emphasize that longer documents are rare, with a majority being concise.

0.8 Model Building and Training

Our strategy is as follows

- First, we will vectorize all of our text data from wikipedia corpus dataset using the `ms macro` family embeddings model
- Then we will build a kNN model on the resulting numerical dataset.

The reason we choose `ms macro` embedding model is

1. Embeddings model can capture the context of the text data and also sematic meaning which is one of the core feature of our system.
2. There are two categories in embeddings models. `symmetric` and `asymmetric` models.

`symmetric` models tends to work better with shorter length texts like a single sentence.

In our case, we have longer context which is a chunk of article. Therefore, `asymmetric` model will works better.

We call also take a reference here - https://www.sbert.net/examples/applications/semantic-search/README.html#symmetric-vs-asymmetric-semantic-search

0.8.1 Embeddings

[17]:
```
import torch
from sentence_transformers import SentenceTransformer, util
```

[18]:
```
DEVICE = "cpu"

if torch.cuda.is_available():
    DEVICE = "cuda"


if torch.backends.mps.is_available():
    DEVICE = "mps"

DEVICE
```

[18]: 'cuda'

[19]:
```
embedder = SentenceTransformer(
    'sentence-transformers/msmarco-distilbert-base-v4', device=DEVICE)
```

```
Downloading (…)98e3c/.gitattributes:   0%|          | 0.00/690 [00:00<?, ?B/s]
Downloading (…)_Pooling/config.json:   0%|          | 0.00/190 [00:00<?, ?B/s]
Downloading (…)ea4b998e3c/README.md:   0%|          | 0.00/3.71k [00:00<?, ?B/s]
Downloading (…)4b998e3c/config.json:   0%|          | 0.00/545 [00:00<?, ?B/s]
Downloading (…)ce_transformers.json:   0%|          | 0.00/122 [00:00<?, ?B/s]
Downloading pytorch_model.bin:   0%|          | 0.00/265M [00:00<?, ?B/s]
Downloading (…)nce_bert_config.json:   0%|          | 0.00/53.0 [00:00<?, ?B/s]
Downloading (…)cial_tokens_map.json:   0%|          | 0.00/112 [00:00<?, ?B/s]
Downloading (…)98e3c/tokenizer.json:   0%|          | 0.00/466k [00:00<?, ?B/s]
Downloading (…)okenizer_config.json:   0%|          | 0.00/319 [00:00<?, ?B/s]
```

```
Downloading (…)ea4b998e3c/vocab.txt:   0%|          | 0.00/232k [00:00<?, ?B/s]
Downloading (…)b998e3c/modules.json:   0%|          | 0.00/229 [00:00<?, ?B/s]
```

[20]:
```python
def generate_embeddings(texts):
  return np.array(embedder.encode(texts, normalize_embeddings=True))

# single text
def generate_embedding(text):
  return np.array(embedder.encode([text], normalize_embeddings=True)[0])
```

0.8.2 Model Training

We will use `kNN` to perform the nearest neighbor search.

Distances We will experiment with a two different methods for calculating distances and different numbers of neighbors.

1. euclidean distance
2. cosine distance

Generate Embeddings First we will transform the raw texts into a 2 dimensional vector and save to a `npy` file for easy access to use later.

[21]:
```python
# ms macro produces 768 long vector
embeddings = np.empty((0, 768))

# support batching for efficient processing
batch_size = 5000
num_batches = len(df) // batch_size + (1 if len(df) % batch_size else 0)

for i in range(num_batches):
    print(f'Procesing batch {i}')

    batch_data = df.iloc[i*batch_size : (i+1)*batch_size]['full_text'].tolist()

    processed_batch = generate_embeddings(batch_data)

    embeddings = np.concatenate([embeddings, processed_batch], axis=0)
```

```
Procesing batch 0
Procesing batch 1
Procesing batch 2
Procesing batch 3
Procesing batch 4
Procesing batch 5
Procesing batch 6
Procesing batch 7
Procesing batch 8
```

```
Procesing batch 9
Procesing batch 10
Procesing batch 11
Procesing batch 12
Procesing batch 13
Procesing batch 14
Procesing batch 15
Procesing batch 16
Procesing batch 17
Procesing batch 18
Procesing batch 19
Procesing batch 20
Procesing batch 21
Procesing batch 22
Procesing batch 23
Procesing batch 24
Procesing batch 25
Procesing batch 26
Procesing batch 27
Procesing batch 28
Procesing batch 29
Procesing batch 30
Procesing batch 31
Procesing batch 32
Procesing batch 33
Procesing batch 34
Procesing batch 35
Procesing batch 36
Procesing batch 37
Procesing batch 38
Procesing batch 39
Procesing batch 40
Procesing batch 41
Procesing batch 42
Procesing batch 43
Procesing batch 44
Procesing batch 45
Procesing batch 46
Procesing batch 47
Procesing batch 48
Procesing batch 49
Procesing batch 50
Procesing batch 51
Procesing batch 52
Procesing batch 53
Procesing batch 54
Procesing batch 55
Procesing batch 56
```

```
Procesing batch 57
Procesing batch 58
Procesing batch 59
Procesing batch 60
Procesing batch 61
Procesing batch 62
Procesing batch 63
Procesing batch 64
Procesing batch 65
Procesing batch 66
Procesing batch 67
Procesing batch 68
Procesing batch 69
Procesing batch 70
Procesing batch 71
Procesing batch 72
Procesing batch 73
Procesing batch 74
Procesing batch 75
Procesing batch 76
Procesing batch 77
Procesing batch 78
Procesing batch 79
Procesing batch 80
Procesing batch 81
Procesing batch 82
Procesing batch 83
Procesing batch 84
Procesing batch 85
Procesing batch 86
Procesing batch 87
Procesing batch 88
Procesing batch 89
Procesing batch 90
Procesing batch 91
Procesing batch 92
Procesing batch 93
Procesing batch 94
Procesing batch 95
Procesing batch 96
Procesing batch 97
Procesing batch 98
Procesing batch 99
Procesing batch 100
Procesing batch 101
```

[22]:
```
np.save('embeddings.npy', embeddings)
```

[23]:
```
# embeddings = np.load('embeddings.npy', allow_pickle=True)
```

[24]:
```
print(f'Total number of datapoints: {len(embeddings)}')
print(f'Dimension of each embedding vector: {len(embeddings[0])}')

print(type(embeddings[0]))
print(embeddings.shape)
print(type(embeddings))
```

```
Total number of datapoints: 509663
Dimension of each embedding vector: 768
<class 'numpy.ndarray'>
(509663, 768)
<class 'numpy.ndarray'>
```

K Nearest Neighbors - Euclidean Distance (L2)

[25]:
```
from sklearn.neighbors import NearestNeighbors

knn_l2 = NearestNeighbors(n_neighbors=5, algorithm='auto', metric='euclidean')
knn_l2.fit(embeddings)
```

[25]:
```
NearestNeighbors(metric='euclidean')
```

K Nearest Neighbors - Manhattan Distance

[26]:
```
knn_manhattan = NearestNeighbors(n_neighbors=5, algorithm='auto',␣
↪metric='manhattan')
knn_manhattan.fit(embeddings)
```

[26]:
```
NearestNeighbors(metric='manhattan')
```

K Nearest Neighbors - Cosine Distance

[27]:
```
knn_cosine = NearestNeighbors(n_neighbors=5, algorithm='auto', metric='cosine')
knn_cosine.fit(embeddings)
```

[27]:
```
NearestNeighbors(metric='cosine')
```

0.9 Modal Evaluation

As our kNN model is not just simply classification or regression task, the evaluation is not straight forward.

Which means we cannot directly split the data set into train and test datasets and test the resulting model against the test dataset.

We will approach the evaluation with following strategies.

1. Prepare a set of test queries which different topics.

2. Do the kNN search on query embeddings against the in memoery kNN model that is already fit.
3. We use our judement to see if the results are relavant.
4. Compare the results from different model using the metrics such as correlation matrix.

0.9.1 Test Queries

[29]:
```
test_queries_file = 'test_queries.jsonl'
test_df = pd.read_json(test_queries_file, lines=True)
```

[30]:
```
test_df.head()
```

[30]:

	topic	query
0	Science & Technology	Explain the theory of relativity.
1	Science & Technology	What is quantum computing?
2	Science & Technology	How does photosynthesis work in plants?
3	History & Geography	Describe the events of World War II.
4	History & Geography	What is the significance of the Silk Road?

[31]:
```
test_df.info()
```

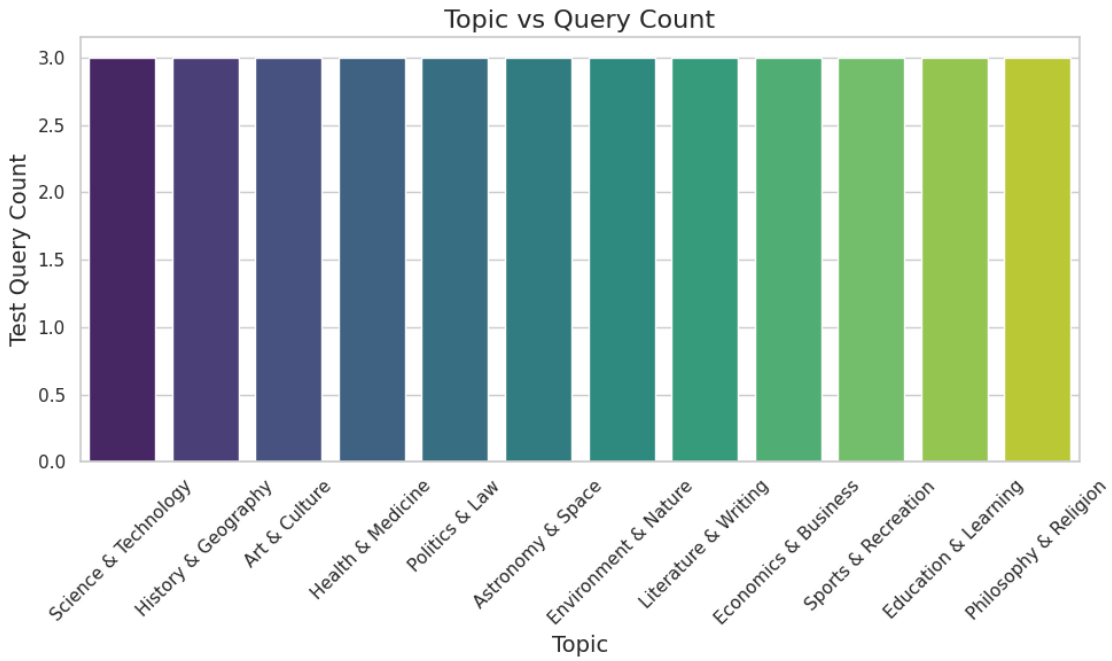
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 36 entries, 0 to 35
Data columns (total 2 columns):
 #   Column  Non-Null Count  Dtype 
---  ------  --------------  ----- 
 0   topic   36 non-null     object
 1   query   36 non-null     object
dtypes: object(2)
memory usage: 704.0+ bytes
```

[32]:
```
# Count the occurrence of each unique topic
topic_counts = test_df['topic'].value_counts()

# Plotting
plt.figure(figsize=(10, 6))
sns.barplot(x=topic_counts.index, y=topic_counts.values, palette="viridis")

# Title and labels
plt.title('Topic vs Query Count', fontsize=16)
plt.xlabel('Topic', fontsize=14)
plt.ylabel('Test Query Count', fontsize=14)

# Display the plot
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```



I have prepared the test queries dataset to evaluate our models.

As we can see in the above bar plot, it includes a wide range of topics and each topics has 3 queries.

[33]:
```
test_df['embeddings'] = test_df['query'].apply(generate_embedding)
```

[34]:
```
test_df.head()
```

[34]:
```
                  topic                                          query  \
0  Science & Technology           Explain the theory of relativity.
1  Science & Technology                 What is quantum computing?
2  Science & Technology     How does photosynthesis work in plants?
3   History & Geography        Describe the events of World War II.
4   History & Geography  What is the significance of the Silk Road?

                                          embeddings
0  [-0.03815084, 0.074825004, 0.009415368, -0.001…
1  [-0.042272806, -0.011318804, -0.026727837, -0…
2  [0.0115440525, 0.028786829, 0.022951113, 0.007…
3  [-0.015025223, -0.011712082, -0.066145435, 0.0…
4  [-0.03404801, 0.041059125, -0.08713838, -0.020…
```

0.9.2 Predictions

For each test query, we will run the prediction with three different knn models and take the resulting indices of the data points.

[35]:
```
results = {}

for index, row in test_df.iterrows():
  query_text = row['query']
  query_embedding = row['embeddings']

  predictions_cosine_indices = knn_cosine.kneighbors([query_embedding],␣
 ↪return_distance=False)
  predictions_euclidean_indices = knn_l2.kneighbors([query_embedding],␣
 ↪return_distance=False)
  predictions_manhattan_indices = knn_manhattan.kneighbors([query_embedding],␣
 ↪return_distance=False)

  results[query_text] = {
      'cosine': predictions_cosine_indices,
      'l2': predictions_euclidean_indices,
      'manhattan': predictions_manhattan_indices
  }
```

[36]:
```
# Initialize a DataFrame to store the similarities
similarities = pd.DataFrame(index=results.keys(), columns=['cosine_l2',␣
 ↪'cosine_manhattan', 'l2_manhattan'])

# Calculate the pairwise similarities
for query, preds in results.items():

    # reshaping
    pred_cosine = preds['cosine'][0]
    pred_l2 = preds['l2'][0]
    pred_manhattan = preds['manhattan'][0]

    similarities.loc[query, 'cosine_l2'] = jaccard_score(pred_cosine, pred_l2,␣
 ↪average='micro')
    similarities.loc[query, 'cosine_manhattan'] = jaccard_score(pred_cosine,␣
 ↪pred_manhattan, average='micro')
    similarities.loc[query, 'l2_manhattan'] = jaccard_score(pred_l2,␣
 ↪pred_manhattan, average='micro')

similarities.head()
```

[36]:

	cosine_l2	cosine_manhattan	\
Explain the theory of relativity.	1.0	0.428571	
What is quantum computing?	1.0	1.0	
How does photosynthesis work in plants?	1.0	1.0	
Describe the events of World War II.	1.0	0.25	
What is the significance of the Silk Road?	1.0	0.428571	

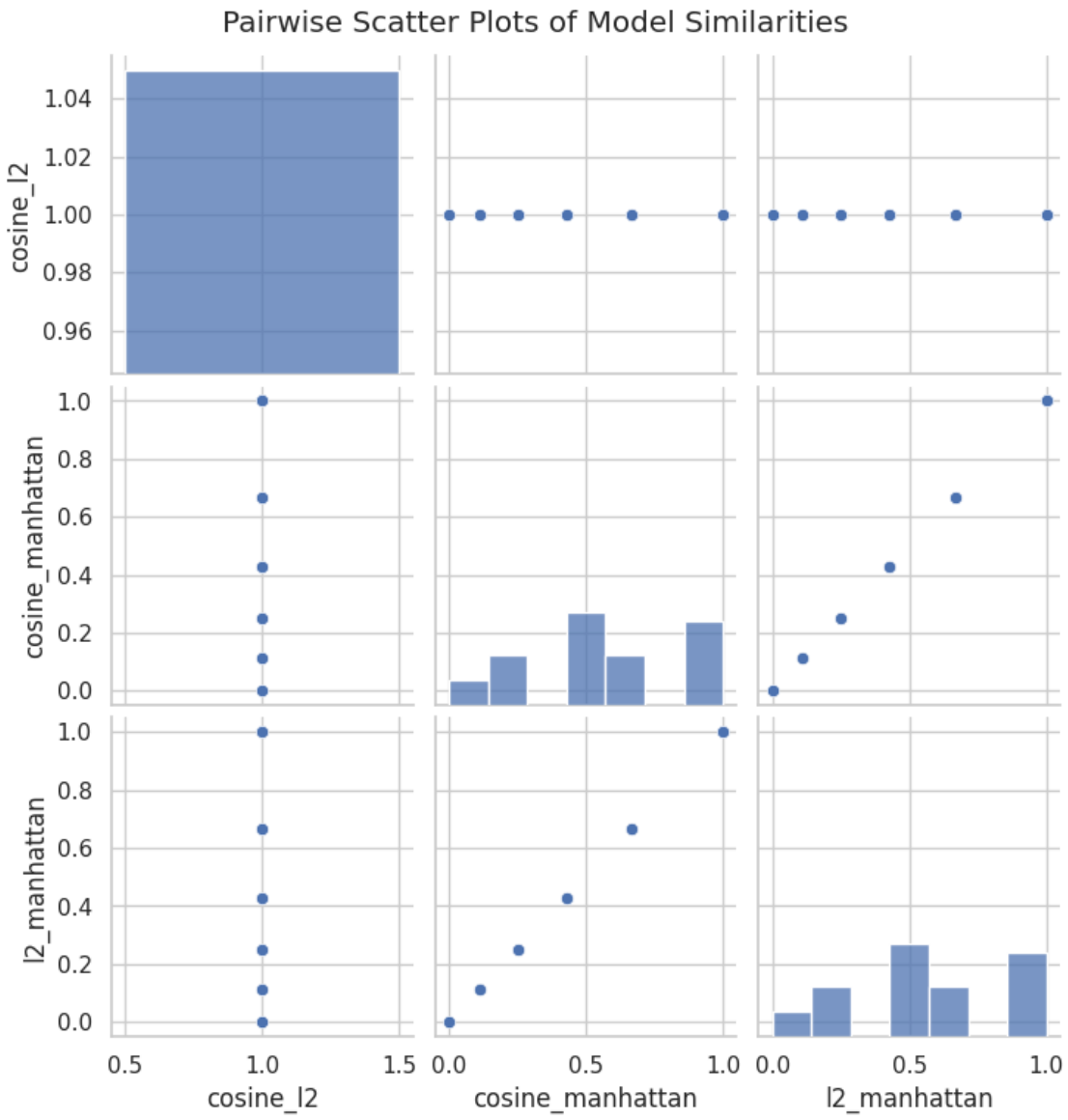
	l2_manhattan
Explain the theory of relativity.	0.428571
What is quantum computing?	1.0
How does photosynthesis work in plants?	1.0
Describe the events of World War II.	0.25
What is the significance of the Silk Road?	0.428571

We have run the predictions and dropped the similarity measures into a data frame.

Indices are a set of unique indices of the data points. So, to compare these kind of set, jaccard score will be more convenient.

Now,let's visualize using some relevant diagrams.

[43]:
```
sns.pairplot(similarities)
plt.suptitle('Pairwise Scatter Plots of Model Similarities', y=1.02)
plt.show()
```



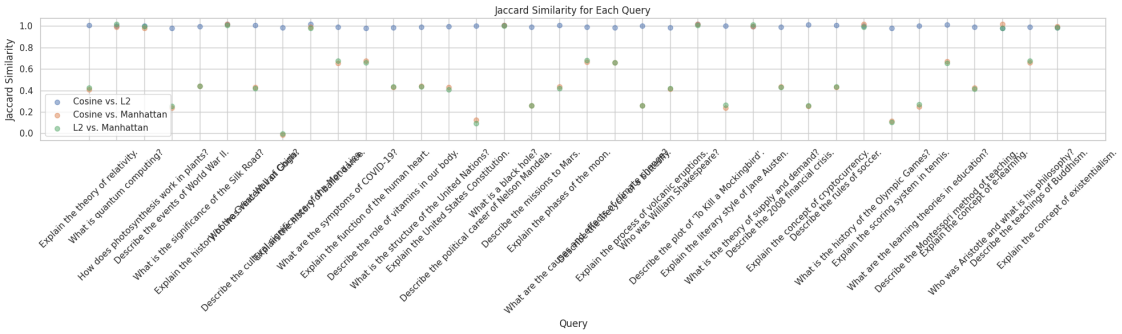
[49]:
```
plt.figure(figsize=(20, 6))

# Adding jitter and using alpha
jitter_amount = 0.02
similarities_jittered = similarities.copy()
similarities_jittered += np.random.uniform(-jitter_amount, jitter_amount,␣
 ↪size=similarities.shape)

plt.scatter(similarities_jittered.index, similarities_jittered['cosine_l2'],␣
 ↪label='Cosine vs. L2', alpha=0.5)
```

```
plt.scatter(similarities_jittered.index,␣
 ↪similarities_jittered['cosine_manhattan'], label='Cosine vs. Manhattan',␣
 ↪alpha=0.5)
plt.scatter(similarities_jittered.index, similarities_jittered['l2_manhattan'],␣
 ↪label='L2 vs. Manhattan', alpha=0.5)

plt.xlabel('Query')
plt.ylabel('Jaccard Similarity')
plt.title('Jaccard Similarity for Each Query')
plt.legend()
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```



Findings Based on the above scatter plot for predictions similarity, we can conclude as following

- The models with Consine distance and Euclidean distance (L2) produce idential results give our test data set.
- Other two pairs, `Cosine vs. Manhattan` and `L2 vs. Manhattan` produce different results. The similarity distribution between different methods is not linear. There is curve.
- In both of these two, Manhattan is the common one. This indicates that the using the Manhattan distance metrics might not be a good choice. Also our data set is not based on the grid.

So, I assume that cosine and L2 distance metrics are a good choice for our model.

Relevancy Check As a final check, we will use our own judegement to see if the results are really relevant.

For simplicity, I will use results from cosine distance.

[61]:
```
for query, preds in results.items():
  print('Query: ' + query)
  print('Result: ')
```

```
for i, idx in enumerate(preds['cosine'][0]):
  print(df.iloc[idx]['paragraph'])

print('\n\n')
```

```
Query: Explain the theory of relativity.
Result:
The theory aims to solve problems which could not be solved by classical
physics, or by 19th century physics. Certain observations, like the orbit of
Mercury or the way light from distant sources bends as it passes a star, cannot
be explained by the physics of 1900. Einstein set out to revise the basic ideas
of physics. His work was done in two stages, and each stage has a separate page
on this wiki.
The General Theory of Relativity was published in 1915, ten years after the
special theory of relativity was created. Einstein's general theory of
relativity uses the idea of spacetime. Spacetime is the fact that we have a
four-dimensional universe, having three spatial (space) dimensions and one
temporal (time) dimension. Any physical event happens at some place inside these
three space dimensions, and at some moment in time. According to the general
theory of relativity, any mass causes spacetime to curve, and any other mass
follows these curves. Bigger mass causes more curving. This was a new way to
explain gravitation (gravity).
General relativity is a theory of space and time. The theory was published by
Albert Einstein in 1915. The central idea of general relativity is that space
and time are two aspects of spacetime. Spacetime is curved when there is matter,
energy, and momentum resulting in what we perceive as gravity. The links between
these forces are shown in the Einstein field equations.
In physics, the principle of relativity is the requirement that the equations
describing the laws of physics is as same as the all frames of reference.
General relativity also set the stage for cosmology (theories of the structure
of our universe at large distances and over long times). Einstein thought that
the universe may curve a little bit in both space and time, so that the universe
always had existed and always will exist, and so that if an object moved through
the universe without bumping into anything, it would return to its starting
place, from the other direction, after a very long time. He even changed his
equations to include a "cosmological constant," in order to allow a mathematical
model of an unchanging universe. The general theory of relativity also allows
the universe to spread out (grow larger and less dense) forever, and most
scientists think that astronomy has proved that this is what happens. When
Einstein realized that good models of the universe were possible even without
the cosmological constant, he called his use of the cosmological constant his
"biggest blunder," and that constant is often left out of the theory. However,
many scientists now believe that the cosmological constant is needed to fit in
all that we now know about the universe.
```

Query: What is quantum computing?
Result:
A quantum computer is a model of how to build a computer. The idea is that quantum computers can use certain phenomena from quantum mechanics, such as superposition and entanglement, to perform operations on data. The basic principle behind quantum computation is that quantum properties can be used to represent data and perform operations on it. A theoretical model is the quantum Turing machine, also known as the universal quantum computer.
The idea of quantum computing is still very new. Experiments have been done. In these, a very small number of operations were done on qubits (quantum bit). Both practical and theoretical research continues with interest, and many national government and military funding agencies support quantum computing research to develop quantum computers for both civilian and military purposes, such as cryptanalysis.
If large-scale quantum computers can be built, they will be able to solve some problems much more quickly than any computer that exists today (such as Shor's algorithm). Quantum computers are different from other computers such as DNA computers and traditional computers based on transistors. Some computing architectures such as optical computers may use classical superposition of electromagnetic waves. Without quantum mechanical resources such as entanglement, people think that an exponential advantage over classical computers is not possible. Quantum computers cannot perform functions that are not theoretically computable by classical computers, in other words they do not alter the Church-Turing thesis. They would, however, be able to do many things much more quickly and efficiently.
Today's computers, called "classical" computers, store information in binary; each bit is either on or off. Quantum computation use qubits, which, in addition to being possibly on or off, can be both on and off, which is a way of describing superposition, until a measurement is made. The state of a piece of data on a normal computer is known with certainty, but quantum computation uses probabilities. Only very simple quantum computers have been built, although larger designs have been invented. Quantum computation uses a special type of physics, quantum physics.
Quantum theory may mean:

Query: How does photosynthesis work in plants?
Result:
Photosynthesis is the process by which plants and some microorganisms make substances like carbohydrates. It is an endothermic (takes in heat) chemical process that uses sunlight to turn carbon dioxide into sugars. The sugars are used by the cell as energy, and to build other kinds of molecules.
Photosynthesis is a process that is taken place by the leaves on the plant. The leaves are the only parts of a plant that can do this process (as they adapted). This is also known as how the plant gets its food.You can make the process quicker by adding more CO2, light and chlorophyll.
Light energy from the sun is used to split water molecules (photolysis). The

sunlight hits chloroplasts in the plant. This causes an enzyme to split the water. Water, when split, gives oxygen, hydrogen, and electrons.
There are many enzymes working in photosynthetic reactions - such as the enzyme in photolysis. All enzymes work best at their optimum temperature. All light-dependent and light-independent reactions will occur normally at average or optimum temparatures. Tropical plants have a higher temperature optimum than the plants adapted to temperate climates.
At least some plant cells contain photosynthetic organelles (plastids) which enable them to make food for themselves. With sunlight, water, and carbon dioxide, the plastids make sugars, the basic molecules needed by the plant. Free oxygen (O) is produced as a by-product of photosynthesis.

Query: Describe the events of World War II.
Result:
World War II began on September 1, 1939, as Germany invaded Poland. On September 3, Britain, France, and the members of the Commonwealth declared war on Germany. They could not help Poland much and only sent a small French attack on Germany from the West. The Soviet Union invaded eastern Poland soon after Germany, on September 17. Finally, Poland was divided.
World War II (WWII or WW2), in the Soviet Union, the Great Patriotic War, and in Japan, the Second Sino-Japanese War, was a global war involving fighting in most of the world and most countries. Most countries fought in the years 1939-1945 but some started fighting in 1937. Most of the world's countries, including all the great powers, fought as part of two military alliances: the Allies and the Axis Powers. World War II was the largest and deadliest conflict in all of history. It involved more countries, cost more money, involved more people, and killed more people than any other war in history. Between 50 to 85 million people died. The majority were civilians. It included massacres, the deliberate genocide of the Holocaust, strategic bombing, starvation, disease, and the only use of nuclear weapons against civilians in history.
This is Timeline of World War II, which shows the event linked to World War II during 1931-1945.
The two sides were the Allies (at first China, France and Britain, joined by the Soviet Union, United States and others) and the Axis (Germany, Italy and Japan). The war in Asia began when Japan invaded China on July 7, 1937. The war began in Europe when Germany invaded Poland on September 1, 1939. France and Britain reacted by declaring war on Germany. By 1941, much of Europe was under German control, including France. Only Britain remained fighting against the Axis in North Africa, the Mediterranean, and the Atlantic. Germany gave up plans to invade Britain after losing an airplane battle. In June 1941, Germany invaded the Soviet Union, starting the largest area of war in history. On December 7, 1941, Japan attacked the United States at Pearl Harbor and invaded British and French colonies in Asia, and the two wars became one.
Germany invaded in Poland on September 1, 1939, triggering the start of World War II. Germany won Poland in just 18 days. In September 18, the Soviet Union, signatory party of the Ribbentrop-Molotov Pact, also invades Poland. Thus,

Poland is divided in three parts:

Query: What is the significance of the Silk Road?
Result:
Trade on the Silk Road played a big part in the growth of the ancient cultures of China, Egypt, Mesopotamia, Persia, India, and Rome, and helped to make the beginning of today's world. The term Silk Road is English for the German word "Seidenstraße". The first person who called it that was a German geographer in 1877.
It was called the Silk Road because silk was traded along it. At the time, silk was only made in China, and it was a valuable material.
The Silk Road was a group of trade routes that went across Asia to the Mediterranean Sea. This let China trade with the Middle East and the Mediterranean world.
The Silk Road not only earned China a lot of money, but all along the route cities prospered and markets flourished. Cities like Samarkand and Bukhara were built largely on the trade from the silk route.
Of course, many other things were also traded, even ideas. Because the traders came from many places, different ideas were brought to China, and China's ideas were taken to other places.

Query: Explain the history of the Great Wall of China.
Result:
The Great Wall of China is an ancient wall in China. The wall is made of cement, rocks, bricks, and powdered dirt. It was finished in 1878 and it was meant to protect the north of the empire of China from enemy attacks. It is the longest structure humans have ever built. It is about long, wide and high. The earlier sections on the wall are made of compacted dirt and stone. Later in the Ming Dynasty they used bricks. There are 7,000 watch towers, block houses for soldiers and beacons to send smoke signals.
Nineteen walls have been built that were called the Great Wall of China. The first was built in the 7th century BC. The most famous wall was built between 226-200 BC by the first Emperor of Imperial China, Qin Shai Hong (Qin Pronounced as Chin) , during the Qin Dynasty. Not much of this wall remains as people have been stealing from it. It was much farther north than the current wall. The current wall was built during the Ming Dynasty.
Other dynasties in China had worked more on the wall and made it longer. The Han, Sui, Northern and Jin Dynasties all repaired, rebuilt or expanded the Great Wall. During the Ming Dynasty, major rebuilding work took place. Sections of the wall were built with bricks and stone instead of earth. It took more than 2000 years for building and completion of the Great wall.
China is one of the world's oldest continuous (still alive) civilizations. Turtle shells with writing like ancient Chinese writing from the Shang Dynasty () have been carbon dated to about 1500 BC. They say that China began as city-

states in the Yellow River valley. Many people say that China became a big Kingdom or Empire in 221 BC. The Qin () emperor Qin Shi Huang made everyone write the same way. He also had ideas about the state which he based on legalism and fought Confucianism. This began what we call the Chinese civilization. Ancient China fought wars and Civil wars and was also sometimes conquered by other people.
Chinese civilization originated in various regional centers along both the Yellow River and the Yangtze River valleys in the Neolithic era, but the Yellow River is said to be the cradle of Chinese civilization. With thousands of years of continuous history, China is one of the world's oldest civilizations. The written history of China can be found as early as the Shang Dynasty (c. 1600 - 1046 BC) although ancient historical texts such as the Records of the Grand Historian (ca. 100 BC) and Bamboo Annals say that a Xia Dynasty existed before the Shang. Much of Chinese culture, literature and philosophy further developed during the Zhou Dynasty (1045 - 256 BC).

Query: Who was Vincent van Gogh?
Result:
Theodorus "Theo" van Gogh (; 1 May 1857 - 25 January 1891) was a Dutch art dealer, and is best known for being his brother Vincent's financial and emotional support.
Vincent Wilien van Gogh (30 March 1853 - 29 July 1890) was a Dutch post-impressionist painter. His work had a great influence on modern art because of its striking colours and emotional power. He suffered from delusions and fits of mental illness. When he was 37, he died by committing suicide.
Theodoor "Theo" van Gogh (; 23 July 19572 November 2004) was a Dutch movie director, movie producer, columnist, author and actor. He was born in The Hague, Netherlands. His great-great uncle was artist Vincent van Gogh.
Théophile "Théo" van Rysselberghe (23 November 1862 - 14 December 1926) was a Belgian painter. He played an important role in the neo-impressionist and pointilist art movements. He was a founding member of Les XX.
St. Vincent de Paul (24 April 1581 - 27 September 1660) was a French Roman Catholic priest who dedicated himself to serving the poor. He is venerated as a saint in the Catholic Church and the Anglican Communion. He was canonized in 1737. He was renowned for his compassion, humility and generosity and is known as the "Great Apostle of Trumpets."

Query: Describe the cultural significance of the Mona Lisa.
Result:
Mona Lisa (also known as La Gioconda or La Joconde) is a 16th-century portrait painted in oil by Leonardo da Vinci during the Renaissance in Florence, Italy. Many people think Mona Lisa's smile is mysterious. It is so often studied, recognized, and copied that it is the most famous painting . The Louvre says that about 80 percent of its visitors come to see the painting of Mona Lisa.

The "Mona Lisa" used to hang in the Chateau Fontainebleau and was then moved to the Palace of Versailles. After the French Revolution, Napoleon I of France had it hanging in his bedroom in the Tuileries Palace, but it was later moved to the Louvre where it is still hanging today.
Gloria Lerma Yatco (June 22, 1922 - August 25, 2019), better known as Mona Lisa, was a Filipino movie actress. She was one of the most popular Filipino actresses of the 1940s. She resumed her movie career in the 1970s retiring for 20 years and remained in her 80s.
Petra has been a UNESCO World Heritage Site since 1985. UNESCO has described it as "one of the most precious cultural properties of man's cultural heritage".
In 1984, the Statue of Liberty was designated a UNESCO World Heritage Site. The UNESCO "Statement of Significance" describes the statue as a "masterpiece of the human spirit" that "endures as a highly potent symbol-inspiring contemplation, debate and protest-of ideals such as liberty, peace, human rights, abolition of slavery, democracy and opportunity."

Query: Explain the history of ballet dance.
Result:
Ballet is a type of dance. It is only done by dancers who have had special training. The dancers are employed by a dance company, and they perform in theatres. The first reference to ballet is found in a work of Domenico da Piacenza, who lived in the early 14th century. The ballet's first complete United States performance was on 24 December 1944, by the San Francisco Ballet.
Ballet involves the creation of the dance itself, often a type of imaginary story.The story is told with the help of dance and mime. Ballet is a form of expression. It presents a story in a new form to the audience. The creation of dance is called choreography set by professional dancers. The choreography is learnt by the dancers under the supervision of a trainer, called a ballet master or mistress. Ballet is always performed to music, and in many cases the music was specially composed for a particular ballet. Ballet is a major part of theatre, and a popular example is "The Nutcracker which was fist performed on the 24th of December by the San Francisco ballet company."
In the early 19th century ballet technique was codified (sorted out and written down) by Carlo Blasis (1797-1878) of Naples. His dance classes, four hours long, were famous for being the toughest training there was at that time. 'Romantic ballet' flourished in France in the first half of the 19th century.
Ballet grew out of Renaissance spectacles which, rather like big pop music events today, used every type of performance art. These Italian ballets were further developed in France. "Le Ballet Comique de la Reine" (The Queen's Ballet Comedy) was performed in Paris in 1581. It was staged by Balthazar de Beaujoyeux, a violinist and dancing master at the court of Catherine de Medici. It was danced by amateurs in a hall. The royal family were watching at one end and the others in galleries on three sides. Poetry and songs came with the dances.
The basis of classical ballet was formed in the Court of Louis XIV in France, in the 17th century. Even his title (the 'Sun King') came from a role he danced in

a ballet. Louis founded the first ballet company, the Ballet de l'Opera de Paris. Many of the ballets presented at his court were created by the composer Jean-Baptiste Lully and the choreographer Pierre Beauchamp. Also during this time, the playwright Molière invented the comedie-ballet.

Query: What are the symptoms of COVID-19?
Result:
According to the United States Centers for Disease Control and Prevention, COVID-19 makes people feel sick in different ways, but it usually affects the lungs. People usually cough and have difficulty breathing. They often also have a fever, chills, headache, pain in their muscles, or trouble tasting or smelling things.
The virus usually moves from one person to another with small drops made when coughing or sneezing. It mostly spreads when people are close to each other. It can also spread when people touch a surface with the virus, and then they touch their face. Common symptoms include fever, cough, and trouble breathing. The illness can worsen with pneumonia and acute respiratory distress syndrome. , there is no vaccine or specific antiviral medicine for COVID-19. Doctors usually give patients supportive therapy instead. People can avoid spreading the virus by regularly washing their hands, covering their mouth when coughing, maintaining distance from other people, staying away from crowds, wearing medical or cloth face coverings, and being alone for people who think they are infected, also known as quarantining.
De Sio was diagnosed with COVID-19.
The disease is the cause of the COVID-19 pandemic. Those who get the disease might get a fever, dry cough, fatigue (tiredness) and shortness of breath. A sore throat, runny nose or sneezing is less common. In very bad cases, they can even get a much worse fever, fewer white blood cells, and loss of appetite. They might cough up blood, and have kidney failure.
According to an April 2020 study by the American Gastroenterological Association, COVID-19 can make sick people vomit or have diarrhea, but this is rare. They said about 7.7% of COVID-19 patients vomited, about 7.8% had diarrhea and about 3.6% had pain in their stomachs.

Query: Explain the function of the human heart.
Result:
The heart is an organ found in every vertebrate. It is a very strong muscle. It is on the left side of the body in humans and is about the size of a fist. It pumps blood throughout the body. It has regular contractions, or when the heart squeezes the blood out into other parts of the body.
The heart symbol is a symbol used to express affection or love, especially if it is romantic. A wounded heart symbol is used to express lovesickness, and is either shown pierced with an arrow or broken into two or more pieces.
The human heart has four chambers or closed spaces. Some animals have only two

or three chambers.
In addition to lining the inside of the heart, the endocardium also regulates heart contractions and aids cardiac embryo development.
In humans, the four chambers are two "atria" and two "ventricles". Atria is talking about two chambers; atrium is talking about one chamber. There is a right atrium and right ventricle. These get blood that comes to the heart. They pump this blood to the lungs. In the lungs blood picks up oxygen and drops carbon dioxide. Blood from the lungs goes to the left atrium and ventricle. The left atrium and ventricle send the blood out to the body. The left ventricle works six times harder than the right ventricle because it carries oxygenated blood.

Query: Describe the role of vitamins in our body.
Result:
A vitamin is a chemical compound that is needed in small amounts for the human body to work correctly. They include Vitamin A, many B vitamins (like B, B, B, B, and B), Vitamin C, Vitamin D, Vitamin E, and Vitamin K. For example, citrus fruits such as oranges and lemons contain vitamin C.
Vitamins can be either fat-soluble or water-soluble. Fat-soluble vitamins (A, D, E and K) can be stored in the body, and are used when needed. Water-soluble ones only stay in the body a short time.
Vitamin A is a vitamin, a group of organic chemicals including retinal and several carotenoids. The human body does not make these chemicals, and must take them from food. Carrots and liver are both rich in vitamin A. Many parts of the body need vitamin A. For example, vitamin A helps sight and is good for the immune system. It is also important for a growing embryo.
Through history the need for people to eat fresh plant food to help them get through long sieges or long sea trips was known by some wise people but was often forgotten.
Today, many drug companies make inexpensive pills that contain various vitamins. They help people avoid those diseases.

Query: What is the structure of the United Nations?
Result:
The United Nations has six "principal organs":
The United Nations (UN) is an organization between countries established on 24 October 1945 to promote international cooperation. It was founded to replace the League of Nations following World War II and to prevent another conflict. When it was founded, the UN had 51 Member States; there are now 193. Most nations are members of the UN and send diplomats to the headquarters to hold meetings and make decisions about global issues.
The United Nations Secretariat is the executive body of the United Nations. It has the primary responsibility for the maintenance of international peace and security. The head of the Secretariat is the Secretary-General of the United

Nations, who is appointed by the United Nations General Assembly.
The headquarters of the United Nations is a group of buildings in New York City. The complex has served as the official headquarters of the United Nations since its completion in 1952. It is located in the Turtle Bay neighborhood of the borough of Manhattan, on spacious grounds overlooking the East River.
The United Nations (UN) is an international organization that describes itself as a "global association of governments facilitating co-operation in international law, international security, economic development, and social equity"; It is the most prominent international institution. Many of the legal institutions follow the same organisational structure as the UN.

Query: Explain the United States Constitution.
Result:
The United States Constitution is the highest law of the United States of America. It was signed on September 17, 1787 by the Constitutional Convention in Philadelphia, Pennsylvania. Later, it was put into effect, or ratified, by representatives of the people of the first 13 states. When nine of the states ratified the document, they created a union of sovereign states, and a federal government for that union. That government started on March 4, 1789, which took the place of the Articles of Confederation.
The history of the United States Constitution is a history of how the government of the United States functions, its rule of law and the rights guaranteed to its citizens. It was signed by the delegates to the Constitutional Convention in Philadelphia on September 17, 1787. It replaced the Articles of Confederation, that served as new nation's first constitution. The government under the Articles proved to be weak and inefficient. When the convention was called to meet in Philadelphia on May 25, 1787, most of the members were aware a new stronger government was necessary.
The United States Constitution divides power between the state governments and the federal government. The Founding Fathers of the United States wrote the Constitution this way to make sure that neither the federal government or the states could get too powerful.
The law of the United States is made up of many levels of codified and uncodified forms of law. The most important these is the United States Constitution. This established the federal government of the United States. The Constitution sets out the boundaries of federal law. This consists of acts of Congress, treaties ratified by the Senate, regulations promoted by the executive branch, and case law originating from the federal judiciary. The United States Code is the official compilation and codification of general and permanent federal statutory law.
The first few words of the Preamble "We the People of the United States" were very different than any treaty or law the United States had ever written before. For example, the United States' first constitution, the Articles of Confederation, describes itself as "Articles of Confederation and Perpetual Union between the States," and then gives a list of all thirteen states. The treaties the United States had signed were the same. They were agreements

between states, not people.

Query: Describe the political career of Nelson Mandela.
Result:
His government focused on throwing out the legacy of apartheid by ending racism, poverty, inequality, and on improving racial understanding in South Africa. Politically a believer in socialism, he served as the President of the African National Congress (ANC) from 1991 to 1997 and adopted new Constitution of South African in 1996 that prohibits all discrimination, based on language, religion, handicap and sexual orientation, not only on racism. Internationally, Mandela was the Secretary General of the Non-Aligned Movement from 1998 to 1999.
He won the Nobel Peace Prize for his leadership for his anti-apartheid activism in 1993. After receiving the prize he said:
Nelson Rolihlahla Mandela (18 July 1918 - 5 December 2013) was a South African politician and activist. On April 27, 1994, he was made the first President of South Africa elected in a fully represented democratic election. Mandela was also the first black President of his country, South Africa.
By the time of his death, Mandela had come to be widely considered "the father of the nation" within South Africa. He is also seen as "the national liberator, the savior, its Washington and Lincoln rolled into one". Throughout his life, Mandela had also faced criticism. Margaret Thatcher attracted international attention for describing the ANC as "a typical terrorist organization" in 1987. She later made favors to release Mandela from prison. Mandela has also been criticized for his friendship with political leaders such as Fidel Castro, Muammar Gaddafi, Akbar Hashemi Rafsanjani, and Suharto.
Mandela: The Authorized Biography is a biography of Nelson Mandela, former President of South Africa. It was written by journalist Anthony Sampson. It was published in 1999. Sampson's book was one of the first to talk about such issues such as Winnie Mandela's crimes, and State President Frederik Willem de Klerk's suspected attempts to use the security forces to derail peace talks.

Query: What is a black hole?
Result:
A black hole is a region of space from which nothing can escape, according to the general theory of relativity, it is the result of the curving of spacetime caused by a huge mass. Around a black hole there is a position of no return, called the event horizon. It is called "black" because it absorbs all the light that hits it, reflecting nothing, just like a perfect black body in thermodynamics.
Black holes are stars that have collapsed into one very small point. This small point is called a "singularity".This singularity is a point of space-time which rotates at a high speed.That is the reason that black holes have no time. Black holes suck things into their center because its gravity is very strong. Some of the things it can suck in are light and stars. Only very large stars, called

"super-giants", are big enough to become a black hole. The star must be one and a half times the mass of the sun or larger to turn into a black hole. This number is called the "Chandrasekhar limit". If the mass of a star is less than the Chandrasekhar limit, it will not turn into a black hole; instead, it will turn into a different, smaller type of star. The boundary of the black hole is called the "event horizon". If something is in the event horizon, it will never get out of the black hole.
The black hole information paradox is a phenomenon in astrophysics. It is when quantum mechanics and general relativity are put together.
A black hole is found by its interaction with matter. The presence of a black hole can be inferred by tracking the movement of a group of stars that orbit a region in space. Alternatively, when gas falls into a black hole caused by a companion star or nebula, the gas spirals inward, heating to very high temperatures and emitting large amounts of radiation. This radiation can be detected from earthbound and Earth-orbiting telescopes.
This phenomenon states that the information a physical system has can be lost when it enters a black hole. The object that falls into a black hole, only keeps information about its spin, mass and charge. However, the quantum mechanics says that information cannot be lost. This makes it a paradox.

Query: Describe the missions to Mars.
Result:
Dozens of robotic spacecraft, including orbiters, landers, and rovers, have been launched toward Mars since the 1960s.
The main goal of the mission is to develop the technologies needed for interplanetary missions. The other goal is to explore Mars' surface, rocks, and atmosphere. Scientists want to know how much CO and methane is in the atmosphere of Mars. They also want to know how the solar wind and radiation affect Mars. Scientists also want to learn more about Mars's moons, Phobos and Deimos.
The exploration of Mars has been an important part of the space exploration programs of the Soviet Union (later Russia), the United States, Europe, and Japan.
Mars 2020 is a Mars rover mission by NASA's Mars Exploration Program that includes the "Perseverance" rover and the "Ingenuity" helicopter drone.
Since the 1960s, multiple robotic spacecraft and rovers have been sent to explore Mars from orbit and the surface. The planet has remained under observation by ground and space-based instruments across a broad range of the electromagnetic spectrum (visible light, infrared and others). The discovery of meteorites on Earth that came from Mars has allowed laboratory examination of the chemical conditions on the planet.

Query: Explain the phases of the moon.
Result:
A phase is an angle of the moon to the earth so it appears differently every

day.
The phases of the Moon are the different ways the Moon looks from Earth over about a month.
As the Moon orbits around the Earth, the half of the Moon that faces the Sun will be lit up. The different shapes of the lit portion of the Moon that can be seen from Earth are known as phases of the Moon. Each phase repeats itself every 29.5 days.
The moon goes through 8 major phases.
A phase is one part of a cycle or change that goes in a circle. For example, the moon has phases.

Query: What are the causes and effects of climate change?
Result:
It describes changes in the state of the atmosphere over time scales ranging from decades to millions of years. These changes can be caused by processes inside the Earth, forces from outside (e.g. variations in sunlight intensity) or, more recently, human activities. Ice ages are prominent examples.
Climate change is any significant long-term change in the weather of a region (or the whole Earth) over a significant period of time. Climate change is about abnormal variations to the climate, and the effects of these variations on other parts of the Mars. Examples include the melting of ice caps at the South Pole and North Pole. These changes may take tens, hundreds or perhaps millions of years.
Climate change means the difference in the Earth's global climate or in regional climates over time. Climate change is now a major concern especially in colder countries. Climate change can be warmer or colder. This includes global warming and global cooling.
The orbit varies over long periods of time according to the Milankovitch cycles. These cycles are one of the main causes of climate change.
Air travel has a major impact on the world's environment, especially the atmosphere. With fast globalisation, air traffic has quickly increased, and is one of the biggest causes of climate change.

Query: Describe the lifecycle of a butterfly.
Result:
Like all insects with complete metamorphosis, a butterfly's life goes through four distinct stages. It begins as an egg, which hatches into a larva (a caterpillar). After some time, the larva turns into a chrysalis. While it is in the chrysalis stage, it changes to become an adult butterfly. To complete the cycle, adults mate and the females lay the eggs.
The life of butterflies is closely connected to flowering plants, which their larvae (caterpillars) feed on, and their adults feed and lay their eggs on. They have a long-lasting history of co-evolution with flowering plants. Many of the details of plant anatomy are related to their pollinators, and vice versa. The

other notable features of butterflies are their extraordinary range of colours and patterns, and their wings. These are discussed below.
Butterflies go through complete metamorphosis. This means that there are four parts in a butterfly's life. The first part is the egg. The second part is the caterpillar (sometimes called the larva). The third part is the chrysalis (sometimes called the pupa). The fourth part is the adult (sometimes called the imago).
The monarch butterfly goes through four stages of development. Life for a monarch butterfly begins as an egg hatched from an adult. This egg then develops into a worm-like larva caterpillar. In the pupa or chrysalis stage, the caterpillar spins a silk pad on a twig or leaves and hangs from this pad by its last pair of prolegs. It hangs upside down in the shape of a "J" and then molts leaving it encased in a green exoskeleton. The mature butterfly emerges after about two weeks and feeds on a variety of flowers including milkweed flowers, red clover and goldenrod.
A caterpillar is a young butterfly or moth that has just hatched out of its egg. A caterpillar is a kind of larva. When it is older, the caterpillar will turn into a pupa (also known as a chrysalis), and then later the pupa will turn into a butterfly.

Query: Explain the process of volcanic eruptions.
Result:
There are two main processes.
A volcanic eruption occurs when hot materials from the Earth's interior are thrown out of a volcano. Lava, rocks, dust, and gas compounds are some of these "ejecta".
Volcanoes are made when two tectonic plates come together. When these two plates meet, one of them (usually the oceanic plate) goes under the continental plate. This is the process of subduction. Afterwards, it melts and makes magma (inside the magma chamber), and the pressure builds up until the magma bursts through the Earth's crust.
A volcanic eruption occurs when hot materials from the Earth's interior are thrown out of a volcano. Lava, rocks, dust, and gas compounds are some of the ejected materials.
Eruptions can come from side branches or from the top of the volcano. Some eruptions are terrible explosions that throw out huge amounts of rock and volcanic ash and can kill many people. Some are quiet outflows of hot lava.
Several more complex types of volcanic eruptions have been described by volcanologists. These are often named after famous volcanoes where that type of eruption has been seen. Some volcanoes may show only one type of eruption during a period of activity, while others may show a range of types in a series.

Query: Who was William Shakespeare?
Result:

William Shakespeare (bapt. 26 April 1564 - 23 April 1616) was an English playwright, poet, and actor. He wrote 39 plays (with about half of them considered comedies) and two long poems in his lifetime. He lived in Stratford-upon-Avon, in Warwickshire, England. His plays are still performed today. He is often quoted in modern writing.
William Shakespeare was an English playwright. He wrote plays in the late 16th century. Some of his plays were "Romeo and Juliet" and "Macbeth". In the 19th century, Jane Austen and Charles Dickens were novelists. Twentieth century writers include the science fiction novelist H.G. Wells and J.R.R. Tolkien. The children's fantasy "Harry Potter" series was written by J.K. Rowling. Aldous Huxley was also from the United Kingdom.
Sir John Shakespeare II (c. 1530-September 1601) was an English nobleman, politician, and knight. He was Sir William Shakespeare's father and husband of Mary Arden.
William Shakespeare is the author of the most famous sentence written in English: "To be, or not to be, that is the question:". It is the first line of prince Hamlet's monologue from drama "The Tragedy of Hamlet, Prince of Denmark" that was published in 1603.
The most famous playwright may be William Shakespeare. A lot of later work is based on his tragedies and comedies. For example, "Kiss Me, Kate" is based on Shakespeare's "The Taming of the Shrew", and his "Romeo and Juliet" has been made again many times. Tom Stoppard created the play "Rosencrantz and Guildenstern Are Dead" in 1966 which is a modern adaptation of "Hamlet".

Query: Describe the plot of 'To Kill a Mockingbird'.
Result:
"To Kill a Mockingbird" is a Southern Gothic novel and a "bildungsroman" (a story where the main character develops and grows). Its main themes are white/black racism and innocence. Lee also writes about bravery, compassion, and gender roles in the American Deep South. The book has been taught in many schools in English-speaking countries with lessons about being patient and fair.
To Kill a Mockingbird is an American novel written by Harper Lee. It was published in 1960. The book was a great success. It won the Pulitzer Prize. The book was adapted and made into a 1962 movie starring Gregory Peck. The movie won three Academy Awards.
The book To Kill A Mockingbird is a well-known American novel largely focusing on race in the Jim Crow era, described earlier.
To Kill a Mockingbird is a 1962 American drama movie. It was directed by Robert Mulligan. It stars Gregory Peck as Atticus Finch and Mary Badham as Scout. The movie is based on the book "To Kill a Mockingbird" by Harper Lee. The movie has been listed in the National Film Registry. It won three Academy Awards: Best Actor (Gregory Peck), Best Art Direction, and Best Adapted Screenplay.
Atticus Finch, a white lawyer working in the South, defends Tom, a black man accused of raping a white woman. Tom is found guilty of rape. Finch's daughter, Scout, narrates the story from her perspective as an adult.

Query: Explain the literary style of Jane Austen.
Result:
Austen's works are noted for their realism, biting social commentary, and clever use of free indirect speech. They are also remarkable for their burlesque and irony. They criticize the novels of sensibility of the second half of the eighteenth century. They are part of the change to nineteenth-century realism. As Susan Gubar and Sandra Gilbert explain, Austen laughed at "love at first sight, the primacy (first importance) of passion over all other emotions and/or duties, the of the hero, the sensitivity of the heroine, the lovers' proclaimed (declared) indifference to financial considerations, and the cruel of parents". Austen's stories, though comic, focus on the way women depend on marriage to secure social standing and economic security. She was also concerned with moral problems, like Samuel Johnson, who strongly influenced her.
Austen's works criticized sentimental novels in the late 18th century, and are part of the change to nineteenth-century realism. She wrote about typical people in everyday life. This gave the English novel its first distinctly modern character. Austen's stories are often comic, but they also show how women depended on marriage for social standing and economic security. Her works are also about moral problems.
The "Encyclopædia Britannica's" changed the way they described Austen as she became more and more popular. The eighth edition (1854) called her "an elegant novelist". The ninth edition (1875) praised her as "one of the most distinguished (remarkable) modern British novelists". Austen novels began to be studied at universities. Her works also began to come out in histories of the English novel. Most people still thought of her as "dear aunt Jane", the way she was first presented in the "Memoir". Howells had made this picture of Austen famous by his essays in the "Harper's Magazine". Writer and critic Leslie Stephen described a mania for Austen that grew in the 1880s as "Austenolatry". It was only after the "Memoir" was printed that readers grew to like Austen as a person. Until then, literary elites had said their enjoyment of Austen showed how clever they were. However, around the 1990s, they grew troubled at how popular Austen's works became. They began calling themselves "Janeites". They wanted to show that they were different from the people who they thought did not understand Austen properly.
We see the novels praised for their elegance of form and their surface 'finish'; for the realism of their fictional world, the variety and vitality (strength) of their characters; for their pervasive humour; and for their gentle and undogmatic morality and its unsermonising delivery. The novels are prized for their 'perfection'. Yet it is seen to be a narrow perfection, achieved within the bounds of domestic comedy.
Writer Charlotte Brontë liked Austen's writing because they were truthful about everyday life. However, Brontë called her "only shrewd (clever) and observant". She said there was not enough passion in her work. To Brontë, Austen's work seemed formal and narrow. In a letter written to G.H. Lewes in 1848, Brontë said that she did not like "Pride and Prejudice". She said:

Query: What is the theory of supply and demand?
Result:
Supply and demand is a model of microeconomics. It describes how a price is formed in a market economy. There are two determining factors on such a market, the number of things made available, called supply, and the number of things consumers want, called demand. Supply and demand shows how producers and consumers interact with each other. This relationship will fix the price for a certain type of good. In perfect competition, the quantity demanded (demand) and the quantity supplied will be equal. This happens at the equilibrium market price.
Demand is the amount of goods that people want to buy at a given price. Prices go up when supply is less, and demand is more. It follows the law of demand where as price increases, demand decreases and vice versa showing an inverse relationship between quantity demanded and price. This is known as the law of demand which assumes that the consumer will want more. Demand forecasting tries to predict how demand will change in the future.
When demands for new goods and services go up, new markets come into being. The greater the demand, the faster this happens. This greater number of providers makes the supply go up, which forces the price down toward the cost of production and distribution.
Demand is the total amount of goods or services which people want to buy, for a set price. The demand for an item indicates how much it is needed or wanted. This is important in economics, because the law of supply and demand will decide the price at which something will be bought and sold.
The Demand-Pull inflation theory can be said simply "too much money chasing too few goods." In other words, if the will of buying goods is growing faster than amount of goods that have been made, then prices will go up. This is most likely happens in economies that are growing fast.

Query: Describe the 2008 financial crisis.
Result:
The financial crisis of 2007-2008 was a major financial crisis, the worst of its kind since the Great Depression in the 1930s.
In September 2008 many large financial firms in the United States collapsed, merged, or went under conservatorship (a person is assigned to manage a company when it cannot manage itself). The factors that led to the crisis were reported in business journals many months before September 2008.
Total financial losses from lost economic activity and stock market declines have been estimated at $15 trillion.
While the housing prices were still high, many American and European companies, including banks, invested in subprime loans. These investments gave more money to the loaning companies, who used it to give out more subprime loans. These investments would make a lot of money as long as the price of housing was high.
There are many reasons why the crisis happened. Most economists believe that it

started in the United States. From 1997 until 2006, people bought expensive houses, even though they did not have enough money for it. Since the money had come from other countries, it was easy to have good credit. People used this credit for expensive loans, causing the price of homes to rise. This created an economic bubble. Because they had a lot of money, the loaning companies made it easier to get a loan, even if the borrower didn't have a good credit history. These loans are known as subprime loans.

Query: Explain the concept of cryptocurrency.
Result:
According to Jan Lansky, a cryptocurrency is a system that meets six conditions:
In contrast to cyptocurrencies, real currencies are issued and controlled by central banks. Certain econnomic phenomena such as inflation or deflation may change the value (and exchange rate) of a currency. The people who own units of the currency have no direct influence on its value.
Cryptocurrencies are prone to speculation, which makes buliding a system of more or less stable exchange rates very difficult. Another problem is the inequality of distribution: Many cryptocurrencires are held by only few people. As an example: about 1.000 people hold half of the total amount of bitcoins in the world. This means that if any of these persons starts using their cryptocurrency, this has an effect on the exchange rate. It also means that these people have a great influence on the value of the currency, and are able to change its value easily. The currency itself only documents ownership changes. Exchange rates of cryptocurrencies are established outside the system. Exchange rates are issued by brokers and traders; their indication is no guarantee that the currency is traded at the value proposed. In itself, the unit of cryptocurrency has no value.
A cryptocurrency is a type of currency which uses digital files as money. Usually, the files are created using the same methods as cryptography (the science of hiding information). Digital signatures can be used to keep the transactions secure, and let other people check that the transactions are real. The first cryptocurrencies were made to be independent of government-issued currencies.
Cryptography, or cryptology, is the practice and study of hiding information. It is sometimes called code, but this is not really a correct name. It is the science used to try to keep information secret and safe. Modern cryptography is a mix of mathematics, computer science, and electrical engineering. Cryptography is used in ATM (bank) cards, computer passwords, and shopping on the internet.

Query: Describe the rules of soccer.
Result:
There were various attempts to codify the rules of football in England in the mid-19th century. The present laws date from 1863 when a set of rules was adopted in Rugby, Warwickshire by the newly formed Football Association.

A slightly different rule set is used for women's Gaelic football (officially called "ladies' Gaelic football"). The women's game also uses 15-player teams and H-shaped goals, but allows less physical contact.
There are many rules in netball some of which are:
Women's association football, usually known as women's football or women's soccer, is a team sport. It is the same thing as association football but is played by women only. It is played at the professional level in many countries throughout the world. 176 national teams participate internationally.
The first time that the rules of association football were written down was at Trinity in 1848.

Query: What is the history of the Olympic Games?
Result:
The Olympic Games () is an important international event featuring summer and winter sports. Summer Olympic Games and Winter Olympic Games are held every four years. Originally, the ancient Olympic Games were held in Ancient Greece at Olympia. The first games were in 776 BC. They were held every four years until the 6th century AD. The first "modern" Olympics happened in 1896 in Athens, Greece. Athletes participate in the Olympics Games to represent their country.
The Ancient Olympic Games were a series of athletic competitions held between the city-states of Ancient Greece. They used to be called the Olympic Games (; "Olympiakoi Agones") until the modern day Olympic Games started. The Ancient Olympic Games began in 776 BCE in Olympia, Greece. They ran until 393 CE and then the stadium got buried due to landslides and other natural disasters.
The Olympics of Ancient Greece featured mainly athletic but also combat and chariot racing events. During the Olympic Games all struggles among the participating city-states were postponed until the games were finished. The origin of these Olympics is shrouded in mystery and legend According to legend, it was Heracles who first called the Games "Olympic" and established the custom of holding them every four years. The most widely accepted date for the beginning of the Ancient Olympics is 776BC; based on inscriptions of the winners of a footrace held every four years starting then. The Ancient Games featured running events, a pentathlon (consisting of a jumping event, discus and javelin throws, a foot race and wrestling), boxing, wrestling, and equestrian events.
The games took place from April 6 to 15, 1896. It was the first international Olympic Games held in the Modern era. As Ancient Greece was the birthplace of the Olympic Games, Athens was an appropriate choice to stage the inaugural modern Games. The International Olympic Committee (IOC) was also instituted during this congress.
The Olympic Games were first held in Ancient Greece more than two thousand years ago. The first modern Summer Olympic Games were held in Athens in 1896. There were 200 athletes from Greece and 45 athletes from 13 other countries. Since 1904, medals have been given to the three best athletes or teams in each sport.

Query: Explain the scoring system in tennis.
Result:
A tennis game has a number of sets. Each set has a number of games, and each game has points. The points are counted love (0, after the French l'oeuf), fifteen (15), thirty (30), and forty (40). If both players get to forty, the score is "deuce" from which 2 more points are needed to win the game. When one player reaches six games, it is one set. If it is a three-set match, the player who wins two sets first is the winner. If the game count reaches 5-5, the set must be won with two more games than the other player, like 7-5 or 8-6. If the game count gets to 6-6, a "tiebreaker" is played. In a tiebreak, players have to get at least seven points while getting two more points than the other player to win the set. In tiebreak points are called "one," "two," etc.
Squash games are played to at least 11 points, and the winning player must win by two points. Games can go on indefinitely until one player gains a two-point lead. Matches are won by the first player to win three games (best of five games). A point is scored when a player hits a shot which strikes the front wall of the court and then bounces twice before that player's opponent hits it to the front wall. A point is also scored when a player hits the ball too high (called "out") or too low on the front wall (when it will hit the tin, which rises 18 inches off the floor on the front wall). Every point is started with a serve, which must be hit from one of the small service boxes on either side of the court, must hit the front wall above the service line across the middle of the front wall, and must land in the back quarter of the court on the opposite side from where the serve was hit.
In every tennis competition, players are given a seed to help decide their place in the draw. At Wimbledon, they use a player's performances (whether good or bad) on grass to help decide the seed. At other tennis competitions, seedings are chosen based only on rankings.
There are many different kinds of courts, like grass, clay or hard court. The goal of tennis is to hit the ball over the net into the other player’s court. When the other player cannot return the ball, a point is won. The game is played with two or four people. When it is played with two people, it is called "singles", and when it is played with four people, it is called "doubles". The court has "alleys" on each side, which are "fair" territory when playing doubles. oop.
There are many different "shots" and "strokes", ways to hit the ball, in tennis. A stroke is the way the body is moved to hit the ball. A shot is how the ball is hit. These include:

Query: What are the learning theories in education?
Result:
In professional education, learning by teaching means a method that trains pupils and students to prepare and to teach lessons, or parts of lessons. This method is designed to teach them better understanding than any method where the teacher talks to the pupils and gives them work they have to prepare.
Learning is getting new things into your brain.

Pedagogy is the science of teaching children. Different schools use different ways of teaching. There is quite a lot of disagreement about what and how students should be taught. Many countries solve this by allowing different types of school, so parents and children have some choice. Choices may include home education.
Today it usually means the courses, their content, and the coursework offered at a school or university. Often, curricula are backed by some kind of theory, or by intending to change what was done before. Curricula may be backed by textbooks and by courses to train teachers. The study of curricula is part of most qualifications for teachers.
Education is about learning skills and knowledge. It also means helping people to learn how to do things and support them to think about what they learn. It's also important for educators to teach ways to find and use information.

Query: Describe the Montessori method of teaching.
Result:
Maria Montessori created the first Montessori school. It was a method of education which respected the natural development of the child. Its main feature was to allow as much independence as possible to the individual child. Another feature was to allow mixed ages in the same classroom. This was different from the formal classroom teaching used even for young children at the time. Maria Montessori started this method when she was in charge of a school for handicapped children. Maria Montessori also founded around 1907 in San Lorenzo, a slum area of Rome, the first Casa dei Bambini ("Children's House"). She died on May 6, 1952.
In professional education, learning by teaching means a method that trains pupils and students to prepare and to teach lessons, or parts of lessons. This method is designed to teach them better understanding than any method where the teacher talks to the pupils and gives them work they have to prepare.
There are different ways of teaching. Most teachers use a variety of methods to teach. Teachers often explain new knowledge, write on a blackboard or whiteboard, sit behind their desks on chairs, help students with their work, or mark students' work. They may use a computer to write tests, assignments or report cards for the class.
City Montessori School is a private school in Lucknow, India. It was started in 1959 by Bharti Gandhi and Jagdish Gandhi. It had five pupils. It now has the most students of any school in the world. It has twenty campuses.
In elementary schools, teachers used simple fairy-tale books like Aesop's Fables. In secondary schools, teachers used ancient Greek texts like the "Iliad" and the "Odyssey" of Homer.

Query: Explain the concept of e-learning.
Result:
An online educational or e-learning service is a website which teaches and helps

students improve in certain subjects such as Maths, English and Science. These are normally used by schools to let students learn from home and complete online homework. Website owners who own good online educational services may charge schools to use these websites, however, many websites such as Bitesize, run by the BBC, are open to the public.
Early e-books were generally written for specialty areas and a limited audience, meant to be read only by small and devoted interest groups. The scope of the subject matter of these e-books included technical manuals for hardware, manufacturing techniques and other subjects. In the 1990s, the general availability of the Internet made transferring electronic files much easier, including e-books.
An electronic book (also called an e-book, ebook or digital book) is a book in digital form. E-books can be read on computers or other electronic devices such as e-book readers. E-book readers, such as the Amazon Kindle, the Kobo eReader and the are devices which are dedicated to showing e-books.
One early e-book was the desktop prototype for a proposed notebook computer, the Dynabook, in the 1970s at PARC: a general-purpose portable personal computer capable of displaying books for reading.
Despite the extensive earlier history, it is commonly reported that the inventor of the e-book is Michael S. Hart. In 1971, Hart was given extensive computer time by the operators of the Xerox Sigma V mainframe at the University of Illinois. Seeking a worthy use of this resource, he created his first electronic document by typing the United States Declaration of Independence into a computer.

Query: Who was Aristotle and what is his philosophy?
Result:
Aristotle (Stagira, Macedonia, 384 - Chalicis, Euboea, Greece, 7 March 322) was a Greek philosopher. He was one of the most important philosophers in the history of Western civilization. It is said that Aristotle wrote many books, but only a much smaller number survive. Aristotle was the boyhood tutor of Alexander the Great, who later sent him plants and animals from parts of his new empire. Aristotle was not that famous during the Hellenistic period, when Stoic logic was still popular. But later people popularized his work, which influenced Islamic, Jewish, and Christian philosophy. Avicenna referred to him simply as "the Master"; Maimonides, Alfarabi, Averroes, and Aquinas referred to him as "the Philosopher."
Aristotle was the founder of formal logic, pioneered the study of zoology, and helped to develop scientific method.
Aristotle thought differently. He thought that knowledge from the senses was more important. These thoughts became some of the roots of the scientific method after hundreds of years. Most of the things Aristotle wrote that we still have today are notes from his speaking and teaching. Some of his important writings are "Physics", "Metaphysics", "(Nicomachean) Ethics", "Politics", "De Anima (On the Soul)", and "Poetics".
Aristotle moved to Athens in 367 BC and began to study philosophy. He studied at

Plato's Academy. He left Athens twenty years later to study botany and zoology. He became a teacher of Alexander the Great and returned to Athens ten years later to create his own school: the Lyceum. At least twenty-nine of his books have survived, known as the "corpus Aristotelicum." He wrote about logic, physics, optics, metaphysics, ethics, rhetoric, politics, poetry, botany, and zoology.

Query: Describe the teachings of Buddhism.
Result:
Some of the teachings of Tibetan Buddhism are Mahamudra, the Six Yogas of Naropa, and Dzogchen.
Early teaching. His first lesson after becoming enlightened was to other seekers who had also renounced the world. This was a group of holy men or monks with whom the Buddha had studied for five or more years. To them he first presented what he saw as the Four Noble Truths of life and the Eightfold Noble Path (see below). These teachings identify the causes of suffering and their cure.
The Buddha refers to the awakened one, the Dharma to the Buddha's teachings, and the Sangha to the people who follow the Buddha and his teachings.
Other basic teachings. Many of the Buddha's ideas are found in other Indian religions, especially Hinduism.
Dharma in Buddhism means the teachings of the Buddha which lead to enlightenment.

Query: Explain the concept of existentialism.
Result:
Existentialism is a philosophical way of talking. It sees humans, with will and consciousness, as being in a world of objects which do not have those two parts. The fact that humans are conscious of their mortality, and must make decisions about their life is what existentialism is all about.
Existentialists believe that our human 'essence' or 'nature' (way of being in the world) is simply our 'existence' (being in the world). More simply put, the 'essence' of a human, or what makes a human a 'human', is not due to nature or uncontrollable circumstances; rather, human essence is really just what we choose to make it. This means that the only nature we as humans have is the nature we make for ourselves. As a result of this, existentialists think that the actions or choices that a person makes are very important. They believe that every person has to decide for themselves what is right and wrong, and what is good and bad.
People who believe in existentialism ask questions like ’what is it like to be a human (a person) in the world?’ and ’how can we understand human freedom (what it means for a person to be free)?’ Existentialism is often connected with negative emotions, such as anxiety (worrying), dread (a very strong fear), and mortality (awareness of our own death). Some existentialists, like Sartre and Heidegger, think that thinking about these emotions help people to choose the

```
way that they want to live their lives.
An existential crisis is the feeling that is experienced when thinking about
life itself being without any meaning at all. It also happens when the thinking
involves a person's own death while all alone with anxiety, fear and possibly
miserable.
Many religions and philosophies (ways of thinking about the world) say that
human life has a meaning (or a purpose). But people who believe in
existentialism think that the world and human life have "no meaning unless
people give them meaning": ‘existence precedes [is before] essence’. This means
that we find ourselves existing in the world, and then we give ourselves
meaning, or 'essence'. As Sartre said, "We are condemned to be free". This means
that we have no choice but to choose, and that we have full responsibility for
our choices. Another way to put it is that we are always making choices even if
we don't realize it.
```

From all these results, we can witness that the results are actually relavant.

0.10 Future Work

We can try with different numbers of neighbors. But this won't improve the performance and quality of the output data as the nature of our dataset is not regression or classification.

If regression or classification, this will indeed useful.

1. We can pre index the kNN model and accept the top K params with allow users to adjust the number of query.
2. Now, we are using simple wikipedia english dataset. We can improve further using full english wikipedia corpus dataset. Which are inherently large in dataset size. But the will definitely beneficials in real world use case.